DOCUMENT RESUME

ED 241 558 TM 840 071

AUTHOR Cleveland, William S.
TITLE Graphs in Scientific Publications.
PUB DATE [83]
NOTE 19p.
PUB TYPE Viewpoints (120) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Computer Programs; *Graphs; Guidelines; Research Tools; *Scholarly Journals; Scientific Research; *Surveys; *Use Studies
IDENTIFIERS Science (Journal)

ABSTRACT

Two surveys were carried out to help increase knowledge of current graph usage in science. A detailed analysis of all graphs in one volume of the journal "Science" revealed that 30 percent had errors. Graphs are used more in some disciplines than in others; a survey of 57 journals revealed natural science journals use far more graphs than mathematical or social science journals. Usage studies such as these provide important information for developing four other areas: new graphical methods for data presentation, guidelines, software, and human graphical perception. (A brief appendix contains guidelines for graphs as they pertain to journals and to authors.) (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED241558

Graphs in Scientific Publications

William S. Cleveland

Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

Graphical communication in scientific publications can be improved; a detailed analysis of all graphs in one volume of the journal *Science* revealed that 30% had errors. Graphs are used more in some disciplines than in others; a survey of 57 journals revealed natural science journals use far more graphs than mathematical or social science journals. Usage studies such as these provide important information for developing four other areas: new graphical methods for data presentation, guidelines, software, and human graphical perception.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

W. S. Cleveland

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
  Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TM 840 071

1. INTRODUCTION

Graphs are vital for communication in science. They are an extremely important means for communicating quantitative information in scientific writings such as journal articles and technical reports from universities, industry, and government. The *Journal of Geophysical Research* provides one example since, as we shall see, approximately one-third of the space of the journal is devoted to graphs; this is clearly a substantial commitment to graphical communication.

Graphs, which are analogous to written language, communicate quantitative and categorical information; written language communicates thoughts, ideas, observations, emotions, theories, hypotheses, numbers, and so forth. Our graphical language is used extensively to convey information because it does so very effectively. Quantitative patterns and relationships in data are readily revealed by graphs because of the enormous power of our eye-brain system to perceive geometrical patterns; this system can quickly summarize vast amounts of quantitative information on a graph, perceiving salient features, or can focus on very specific detail. And the recent explosion in computer graphics hardware and software now enables us both to use more graphs and to design new types.

Graphical communication in science needs improvement, as the survey information in this paper demonstrates. Five areas where further research and development could assist this improvement are

(1) Carrying out studies of how graphs are used

(2) Developing new methods for data presentation

(3) Developing guidelines

(4) Studying human graphical perception

(5) Developing software for statistical graphics

The first of the five areas — usage studies — is extremely important for the other four. Through usage studies we find out what methods are in use and thus what software is needed. Through usage studies we can pinpoint common errors and poor designs; guidelines and methods of graphical

presentation should be developed to combat such problems. And studies of usage help shape the decision of what graphical forms should be involved in experiments on human graphical perception.

Wainer (1980) and Tufte (1983) have studied graph usage in the mass media and have found widespread misrepresentation of data and egregious errors. Their work has led to important suggestions on how graphs can be improved in the mass media (Tufte, 1983). The same path should be followed for graphs in science. The blatant errors and misrepresentations of the mass media do not occur so commonly in scientific presentation graphics. However, serious, but different, problems occur in science.

In this paper I will report on two surveys of current graph usage in science. One measured the amount of usage in 57 scientific journals. The design of this survey benefited from a historical survey of two statistics journals carried out by Fienberg (1979). The other survey was much more detailed and intensive. I studied the 377 graphs in the 37 articles and 212 reports of the 1980 Volume 207 of *Science* and recorded types, errors, purposes, unconventional practice, possible methods of improvement, and a number of other variables. The reason for choosing *Science* was that it has a mixture of papers from many different disciplines. In this paper I will report on graph types and errors.

2. DETAILED SURVEY OF *SCIENCE*

One of the tasks in the detailed survey of the 377 graphs in Volume 207 of *Science* was to record four types of what are called "errors", which are classified as follows: (1) Construction — a mistake in the construction of the graph such as tick marks incorrectly spaced, mislabelling, items omitted, and wrong scales. (2) Degraded Image — some aspect of the graph is missing or partially missing due to poor reproduction. (3) Explanation — something on the graph is not explained. (4) Discrimination — items on the graph, such as different symbol types, cannot be easily visually distinguished due to the design or size of the graph.

An example of a construction error is Figure 1 on page 408 of *Science*. The variable on the vertical axis is plotted on a log scale; it was intended that the tick marks be labelled by 10 to an

exponent, but the exponents were omitted so that all tick marks are labelled 10. An example of a severely degraded image is in Figure 1 of page 79. It was intended that a variable (immunoreactivity) be portrayed by shaded areas, but the shading is barely visible due to poor reproduction and the values cannot be seen. An example of an error of explanation is Figure 1 on page 75. Measurements are shown as individual points along 13 different vertical lines corresponding to 13 categories; for each category there is a superposed bar, as on a bar chart, and a single error bar. What the bars and the error bars represent is not explained. An example of poor visual distinguishability is Figure 1 on page 763. On the bottom panel two symbol types, circles and squares, are used to represent two types of observations, but because of overplotting of the symbols, neither the types nor even the number of observations can be fully perceived.

Figure 1 shows the percent of the 377 graphs that had errors; the information is portrayed by number and by type. 30% of the graphs had at least one error. The most frequent type was errors of explanation; many of these were a line or curve superimposed on a plot of data with no explanation of whether the line or curve was drawn by eye, fitted by least squares, or drawn by some other procedure. The error rates shown in Figure 1 seem high particularly when we consider that there are almost certainly errors that I did not detect. And it is likely that *Science* does not have higher error rates than many other journals.

The appendix gives easily implemented guidelines for authors and journals that would, I think, substantially reduce the kinds of errors that were detected. Some of these guidelines were part of a set that I prepared for the journals of the American Statistical Association on behalf of its Committee on Statistical Graphics (American Statistical Association, 1982).

In the detailed survey, graph types were also recorded. A graph was first categorized by the number of variables represented. If the distribution of measurements of a single variable was portrayed, the graph was put in the one-variable group even when the values were grouped by a categorical variable. Figures 1 to 5 of this paper fall in the one-variable group. The one-variable group was further subdivided into bar charts, histograms, point graphs (points plotted along a line), and miscellaneous

If a graph portrayed two paired variables (with numerical values) on an xy graph by graphing one variable against another, it was put into the two-variable group. Sometimes a graph consisted of two or more such xy plots that were either juxtaposed, so that one scale was shared by all panels, or that were superposed; in these cases the graph was still put in the two-variable group. The two-variable group was further subdivided according to the types of variables portrayed. If one variable was time, the graph was put in the "time" group. If one variable was controlled, that is, the values of the variable were specified, say, in an experiment, then the graph was put in the "controlled" group. Since time is in essence usually a controlled variable, the time group is actually a subgroup of the controlled group. If neither variable was controlled, that is, both were just observed, then the graph was put in the "uncontrolled" group.

When a graph attempted to portray the joint structure of three variables, it was put in the three-variable group. Figure 6 of this paper is a three-variable graph. One subgroup, which accounted for almost all such graphs, was statistical maps in which a numerical variable was portrayed (e.g., by symbol type or by contours) as a function of geographical location. No figure in the survey used any of the many displays devised to portray the structure of data in more than three dimensions (Gnanadesikan, 1977; Feinberg, 1979; Tukey and Tukey, 1981; Chambers et al., 1983).

Figure 2 shows the percent of the 377 graphs in the detailed survey of *Science* that fall into each of the three main categories (one-, two-, and three-variable graphs) and into each of the subcategories of the main categories. Thus the percents for the three main categories add to 100% and the percents for the subcategories of a particular category add to the percent for the category. Clearly the most prevalent category is the two-variable one, which accounts for 83% of the graphs; 75% are either in the controlled or in the time category. Thus a majority of the graphs in *Science* show how one measured variable depends on a controlled variable or time.

3. AMOUNT-OF-USAGE SURVEY OF 57 JOURNALS

In the amount-of-usage survey, 57 journals from 14 disciplines were selected. For each journal, 50 articles from the period 1980-1981 were sampled; the number 50 was used since a pilot study

had indicated that standard errors of measured quantities would be sufficiently small to allow inferences about the entire population of articles during this time period. The measurements made on each article that are reported here are the total area of the article, the total area of the graphs in the article, and the number of graphs in the article. The amount of space that could have been used for text had the graph and its legend not been present was measured and then the area that could have been used for text had the legend not been present was measured. The graph area measurement is the first minus the second. Thus graph area is really the amount of text replaced by a graph. A figure was judged to be a graph if it had scales and conveyed quantitative information; however, maps that had no quantitative information other than geographical location were not taken to be graphs.

Figure 3 shows, for each of the 57 journals, the fraction of the total area of the 50 articles devoted to graphs. The largest fraction is that of the *Journal of Geophysical Research*, which has a fractional graph area of 0.310, and the smallest is *The Journal of Social Psychology*, which has a fractional graph area of zero. The median fraction is 0.066, the 75th percentile is .108, and the 25th percentile is 0.032.

Estimates of the standard errors of the fractional graph areas tended to increase with increasing areas; for areas less than 0.025, the standard errors were less than 0.01 and for areas greater than 0.025, standard errors ranged from 0.007 to 0.021. The probabilistic model upon which the standard error calculation was based is that the graph areas and total areas for the 50 articles are a sample of 50 observations from the same bivariate distribution. The fractional graph area then is the sum of the values for the first variable divided by the sum of the values of the second variable. The standard error of the ratio for each journal was computed using Taylor series approximations (Kendall and Stuart, 1977). Standard errors using the bootstrap (Efron, 1982; Efron and Gong, 1983) were computed for *Science* articles and *Science* reports, and there was very close agreement with the Taylor series approximations.

The journals in Figure 3 are arranged according to three categories of discipline. In this grouping *The Annals of Statistics* has been put in the mathematical category since the proving of

theorems is the main goal of most articles in the journal and not the development of statistical methodology nor the analysis of data. It would be unwise to compare individual disciplines since the sample sizes are so small, but it does seem reasonable to draw conclusions about the three broad categories: natural, mathematical, and social sciences. Clearly graph usage is much greater among the natural science journals than among the social science and mathematical science journals.

The lesser use of graphs in the mathematical sciences than in the natural sciences quite likely stems from the journals in mathematics, computing, and statistics having far less observational data to present; they focus mostly on mathematics and methodology. However, many of the social science journals have much data and yet make very little use of graphs.

The fraction of space a journal uses in printing graphs does not measure the effectiveness or the quality of the graphs. It is certainly possible that a journal with less of its space devoted to graphs than another journal might be communicating more effectively with its graphs. But the fraction of space used is a measure of the importance that the journal and the authors attach to graphical communication.

There is large variation in the fractional areas of Figure 3. That is, journals vary substantially in the amount of graphics they employ. Is this due to variation in the numbers of graphs in journals or to variation in the average size of the graphs? To help answer this question the average size, or area, of graphs in each of 50 of the 57 journals was computed; sizes for journals for which the fractional graph area was less than 0.01 were not computed since an accurate estimate of typical graph size for a journal with very few graphs cannot be obtained.

The 50 sizes are shown on a log scale in Figure 4. Base 2 logarithms are used; thus a change of one on the log cm^2 scale means a doubling of values on the cm^2 scale. Figure 5 portrays five of the sizes by squares whose areas are equal to the sizes. The five values are the minimum, lower quartile, median, upper quartile, and the maximum. The ratio of the maximum to the minimum is about 4. The standard errors of the log sizes were estimated using Taylor series approximations (Kendall and Stuart, 1977). The estimates are small; 50% are less than .1, 75% less than .15, and

90% less than .23. Furthermore, the estimated standard errors appear independent of the log sizes.

The measure of fractional graph area presented in Figure 3 is

$$f = \frac{\text{graph area}}{\text{total area}}$$

$$= \frac{\text{graph area}}{\text{number of graphs}} \times \frac{\text{number of graphs}}{\text{total area}}$$

$$= s\,d$$

where s = (graph area)/(number of graphs) is the average size and d = (number of graphs)/(total area) is the average density of occurrence. It is log(s) that is shown in Figure 4. On a log scale we have

$$\log(f) = \log(s) + \log(d) .$$

Figure 6 shows the 3 pairwise scatterplots of the 3 variables on a base 2 log scale; the top panel is $\log(d)$ vs $\log(s)$, the bottom left panel is $\log(f)$ vs. $\log(s)$, and the bottom right panel is $\log(f)$ vs. $\log(d)$. There are 50 journals in Figure 6; these are the same 50 that appeared in Figure 4. On the six scales of the three panels in Figure 6 the number of log units per cm is the same; this is important because of the graphical comparisons we are about to make.

Figure 6 shows two things. The first is that size and density are not strongly related. The second is that variation in density is far more responsible for the variation in the fractional graph areas than is the variation in graph size. We can see this clearly on the two lower panels. On the right one, if two points have the same log density but differ in log fractional area, the vertical distance between them is equal to the difference in their log sizes. On the left one, if two points have the same log size but differ in log fractional area, the vertical distance between them is equal to the difference in their log densities. It is clear that for log density held fixed there is far less variation in log size than there is in log density when log size is held fixed. Thus variation in fractional graph area among the journals is due more to variation in the number of graphs than to variation in average size.

4. DISCUSSION

Statisticians can play an important role, indeed the leading role, in effecting an improvement of graphical communication in science. Since statistics has traditionally dealt with methods for analyzing *and* communicating data, it is natural that ways of analyzing and communicating data with graphs be thought about by statisticians. Recently, three books on statistical graphics have appeared (Chambers et al., 1983; Schmid, 1983; Tufte, 1983). But much more on graphical communication in science needs to be done. What are the areas that appear to need more thought?

One is the assessment and development of graphical methods for data presentation; by methods I mean types of graphs and broad graphical principles. A companion article (Cleveland, 1983) deals with this area. Another area is the development of guidelines to insure clarity of graphs in science — both visual clarity and clarity of explanation. The guidelines in the appendix can be helpful here.

A third important area is studying human *graphical perception*: how people visually decode the quantitative information encoded on a graph. Basic principles of human graphical perception can be used to help develop more effective graphical methods. This is illustrated in the companion article (Cleveland, 1983). The main criterion for judging a graph should be how well people extract the quantitative information portrayed on it. The area of graphical perception combines statistical graphics and visual perceptual psychology to study how people perform this extraction. Almost all past work in this area has consisted of experiments in which subjects look at graphs and record their judgments (for reviews see Feinberg and Franklin, 1975; Kruskal, 1975 and 1982; Cleveland, Harris, and McGill, 1983). But recently, Cleveland and McGill (1983) have developed a theory for one aspect of graphical perception — the relative accuracy of 10 basic perceptual tasks that people employ to visually decode quantitative information encoded on graphs.

A fourth important area is software design. The computer graphics revolution has allowed us to use far more graphs, to get them more easily, and to design new methods that previously would have been impossible or difficult. All of the graphs in this paper were drawn using a pen plotter driven by statistical graphics software on a super-mini computer. Clearly, easy-to-use but yet flexible

software is an important factor in graphical communication in science.

A proper understanding of these four areas depends very much on knowledge of a fifth: how people currently use graphs in science. This fifth area has been the main topic of this paper. Two surveys were carried out to help increase our knowledge of current graph usage in science. One survey showed a large number of errors in one volume of *Science* and another showed a greater use of graphs in the natural sciences than in the mathematical and social sciences. Statisticians, mathematicians, computer scientists, and social scientists need to carefully consider the use of graphs in their disciplines to see whether an important communication tool is being used less than it might be. All scientists need to carefully consider whether an important communication tool is being used less effectively than it might be.

APPENDIX: GUIDELINES FOR GRAPHS

Guidelines for Journals [1] In the instructions to referees ask them to review graphs and to suggest modifications if needed. [2] Inform authors that graphs will be reviewed; in the written information for authors include the guidelines for authors given below. [3] Pay careful attention to reduction and reproduction that graphs must endure. Inform authors about reduction policy so they can design graphs to accommodate it.

Guidelines for Authors [1] Graphs must be visually clear and capable of withstanding reduction. For example, lines must be thick enough, letters must be large enough, and plotting symbols must be large enough and clear enough to accommodate the reduction. A reducing photocopier is often useful in judging how well an original can stand reduction. [2] Graphs must be clearly described. The combined information in the legend and in the text in the body of the paper, should provide a clear and complete description of *everything* that is on the graph. Detailed figure legends can often be of great help to the reader. First describe completely what is graphed in the display, then draw the reader's attention to salient features of the display, and then briefly state the importance of these features. [3] When feasible, put important conclusions into graphical form. Not everyone reads an entire article from beginning to end. When readers skim a paper they are drawn toward graphs.

Try to make the graphs and their legends tell the story of your article. [4] Make the quantitative information that is graphed stand out. Be sure that different items on a graph can be easily visually distinguished. [5] Avoid cluttering graphical displays. For example, too much writing on the plotting region can interfere with the viewer's perception of geometric patterns; put as much of the writing as possible — for example, a key for symbol types — outside of the plotting region, unless you are certain the writing will not interfere. [6] Proofread graphs.

ACKNOWLEDGMENTS

Marylyn McGill of MEM Research, New Providence, N.J. made the measurements in the amount of usage survey and generated the graphs of the paper. To her I am greatly indebted for her substantial care and accuracy. Jon Kettenring, Colin Mallows, Marylyn McGill, Robert McGill, Walter Tapp, and Paul Tukey made many helpful comments on the paper.

REFERENCES

American Statistical Association (1982), "Preparation of Manuscripts for ASA Journals," *Journal of the American Statistical Association* 77, 226-229.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A., (1983), *Graphical Methods for Data Analysis*, Wadsworth International Group (hardcover), Belmont, Calif. and Duxbury Press (softcover), Boston, Mass.

Cleveland, W. S. (1983), "Graphical Methods for Data Presentation: Dot Charts, Full Scale Breaks, and Multi-based Logging," submitted for publication.

Cleveland, W. S., Harris, C. H., and McGill, R. (1983), "Experiments on Quantitative Judgment of Graphs and Maps," *The Bell System Technical Journal*, 62, 1659-1674.

Cleveland, W. S., and McGill, R. (1983), "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods," submitted for publication.

Efron, B. (1982), *The Jackknife, the Bootstrap, and Other Resampling Plans*, SIAM, Philadelphia, Penna.

Efron, B. and Gong, G. (1983), "A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation," *The American Statistician*, 37, 36-48.

Feinberg, B. M. and Franklin, C. A. (1975), *Social Graphics Bibliography*, Bureau of Social Science Research, Washington, D.C.

Fienberg, S. E., (1979), "Graphical Methods in Statistics," *The American Statistician*, 33, 165-178.

Gnanadesikan, R., (1977), *Methods for Statistical Data Analysis of Multivariate Observations*, Wiley, New York.

Kendall, M. G. and Stuart, A. S., (1977), *The Advanced Theory of Statistics*, Vol. 1, 4th ed., Macmillan, New York.

Kruskal, W. H. (1975), "Visions of Maps and Graphs," *Auto-Carto II, Proceedings of the International Symposium on Computer Assisted Cartography*, September 21-27, 1975, edited by J. Kavaliunas, U.S. Bureau of the Census and American Congress on Survey and Mapping, 27-36.

Kruskal, W. H. (1982), "Criteria for Judging Statistical Graphics," *Utilitas Mathematica*, 21B, 283-310.

Schmid, C. F. (1983), *Statistical Graphics*, Wiley, New York.

Tufte, E. R., (1983), *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, Conn.

Tukey, P. A., and Tukey, J. W., (1981), "Graphical Display of Data in Three and Higher Dimensions," *Interpreting Multivariate Data*, edited by V. Barnet, Wiley, New York, Chapters 10, 11, and 12.

Wainer, H., (1980), "Making Newspaper Graphs Fit to Print," *Processing of Visible Language*, Vol. 2, edited by Bouma, Kolers, and Wrolsted, Plenum, New York.

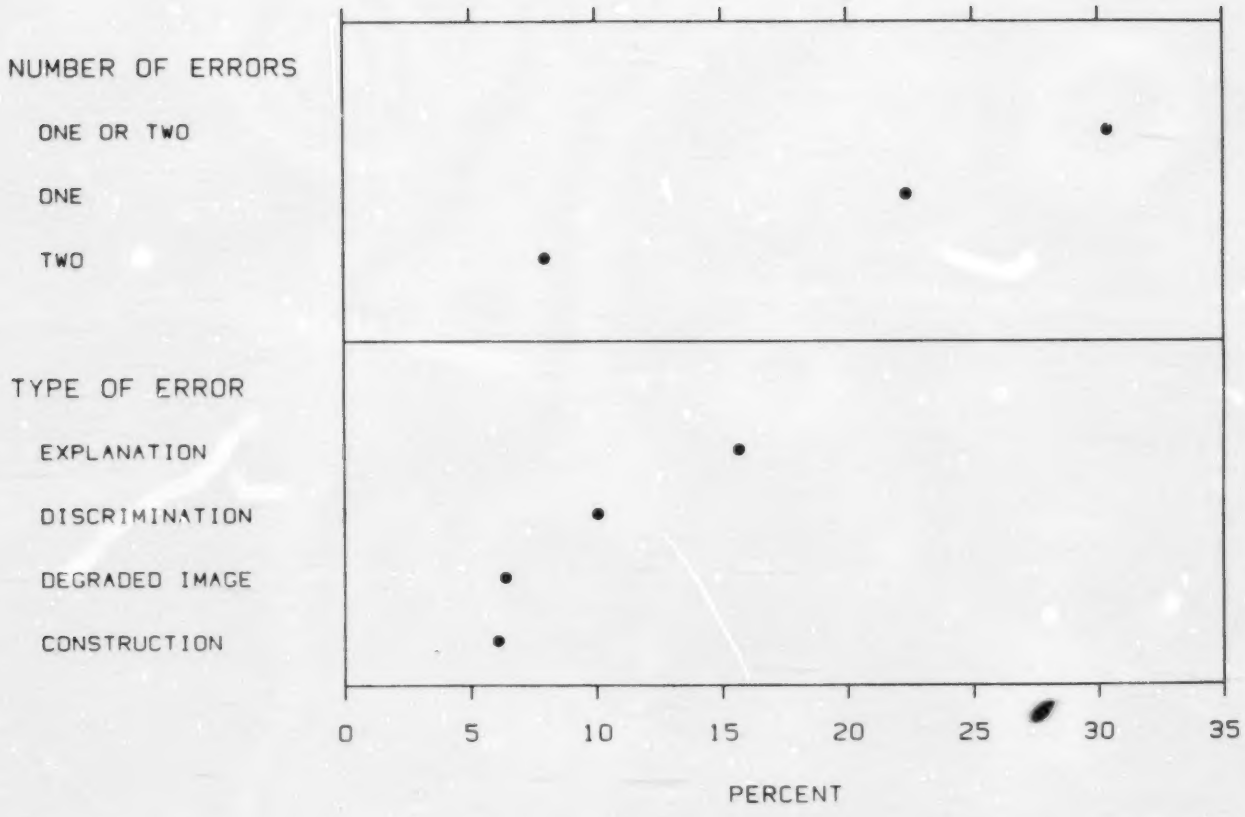

Figure 1. NUMBER AND TYPES OF GRAPH ERRORS IN *SCIENCE*. A detailed analysis of the 377 graphs in the 37 articles and the 212 reports of the 1980 Volume 207 of *Science* was carried out. Four types of errors — deficient explanation, poor reproduction (degraded image), poor visual discrimination, and construction errors — were recorded. The top part of the figure shows the percents of the 377 graphs that had exactly one type of error, exactly two types of errors, and at least one error (no graph had more than two). The bottom part of the figure shows the percent that had each type of error.

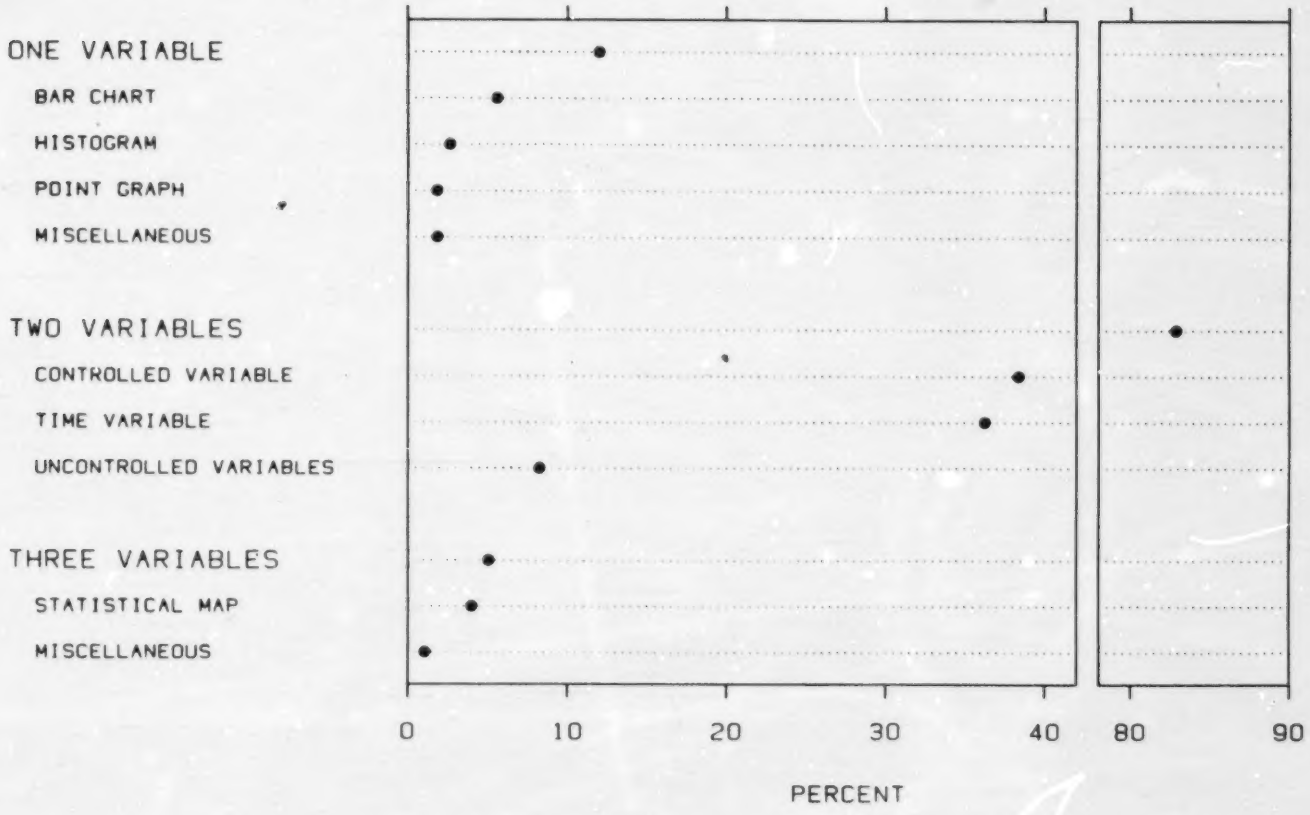

Figure 2. TYPES OF GRAPHS IN *SCIENCE*. In the detailed survey of *Science* each of the 377 graphs was categorized by type. The figure shows the percent in each of three main categories (one, two, and three variables) and in subcategories of each main category. A large majority of the graphs (75%) are two variable xy graphs with one variable either time or a controlled variable.

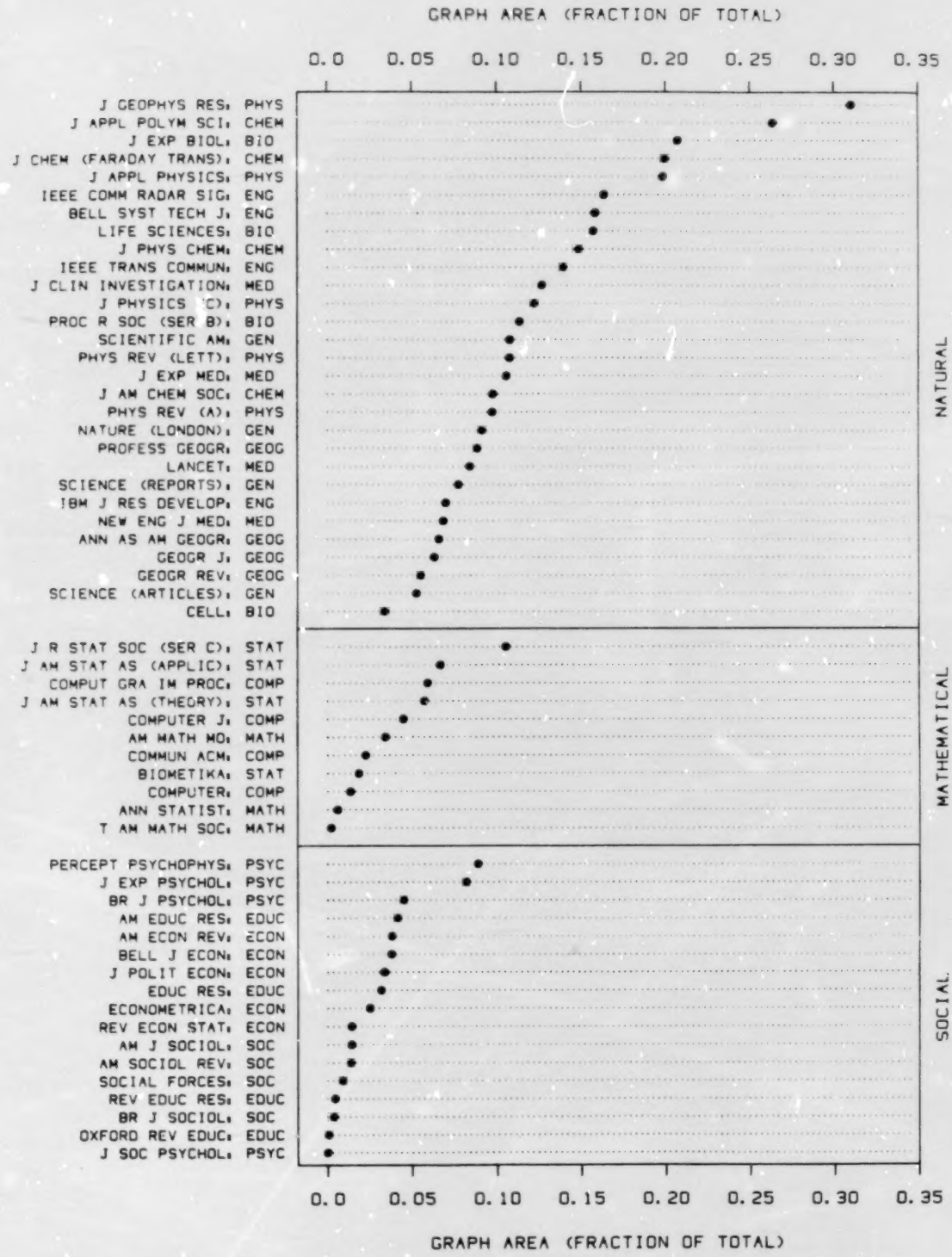

Figure 3. FRACTIONAL GRAPH AREA FOR 57 JOURNALS. Fifty articles published during 1980 and 1981 were sampled from each of 57 scientific journals. The figure shows, for each journal, the fraction of the area of the 50 articles devoted to graphs. A graph is defined to be a figure with scales that communicates quantitative information, and fractional graph area is a measure of the amount of graphical communication in a journal. The journals are grouped according to three categories of scientific disciplines: natural science, mathematical science, and social science. The natural science journals tend to make far greater use of graphs than the mathematical science and social science journals.

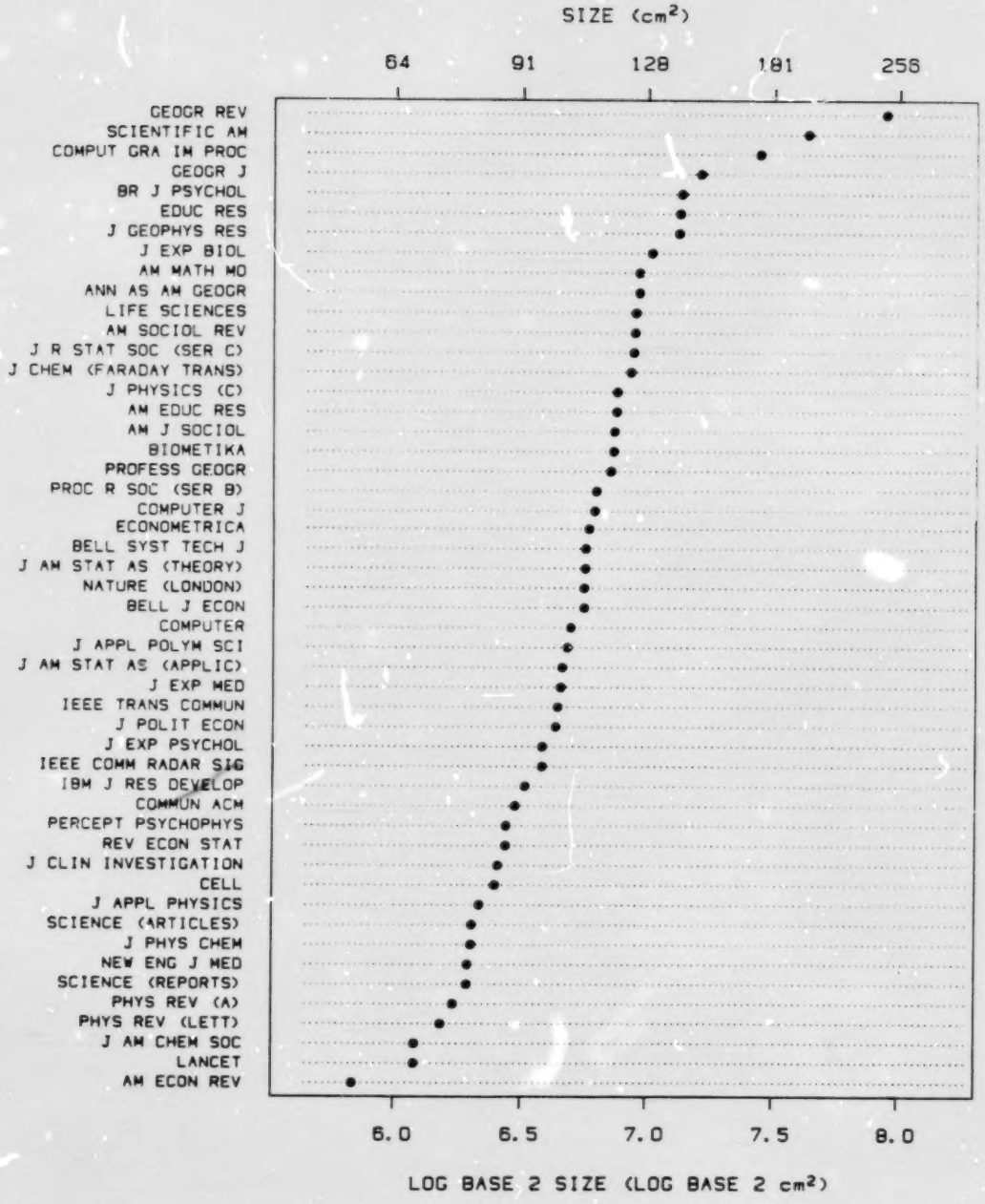

Figure 4. AVERAGE SIZE OF GRAPHS FOR 50 JOURNALS. The logarithms (base 2) of the average areas of graphs in 50 of the 57 scientific journals are shown. The 7 journals omitted are those whose fractional graph areas are less than 0.01. Two scales are portrayed by the upper and lower horizontal axes of the graph. The lower is the base 2 log cm^2 scale and the upper is the cm^2 scale.

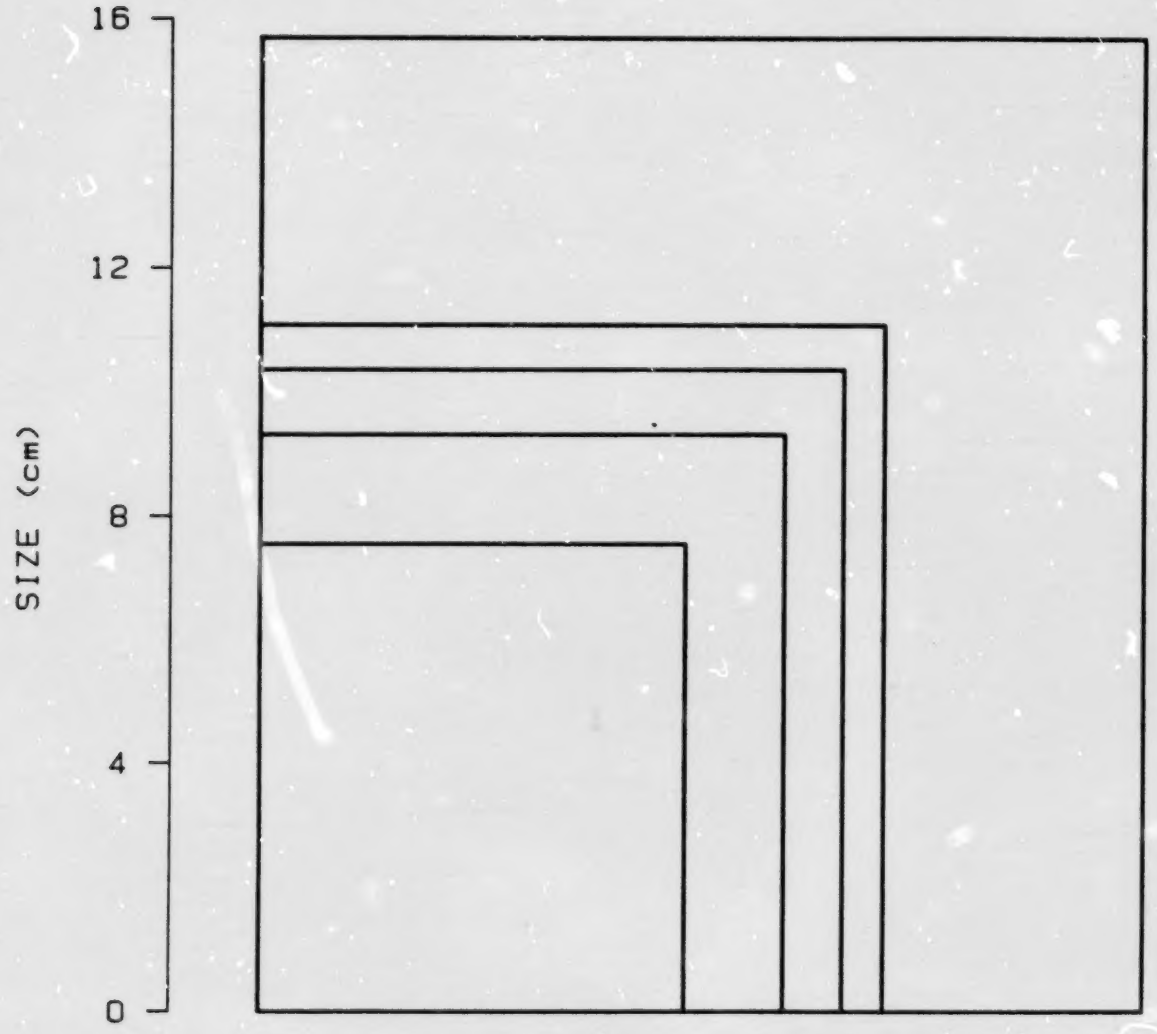

Figure 5. FIVE AVERAGE SIZES. The areas of the squares are equal to 5 of the Figure 4 average graph areas: the smallest (57 cm^2), the 25th percentile (87 cm^2), the 50th percentile (107 cm^2), the 75th precentile (123 cm^2), and the maximum (247 cm^2).

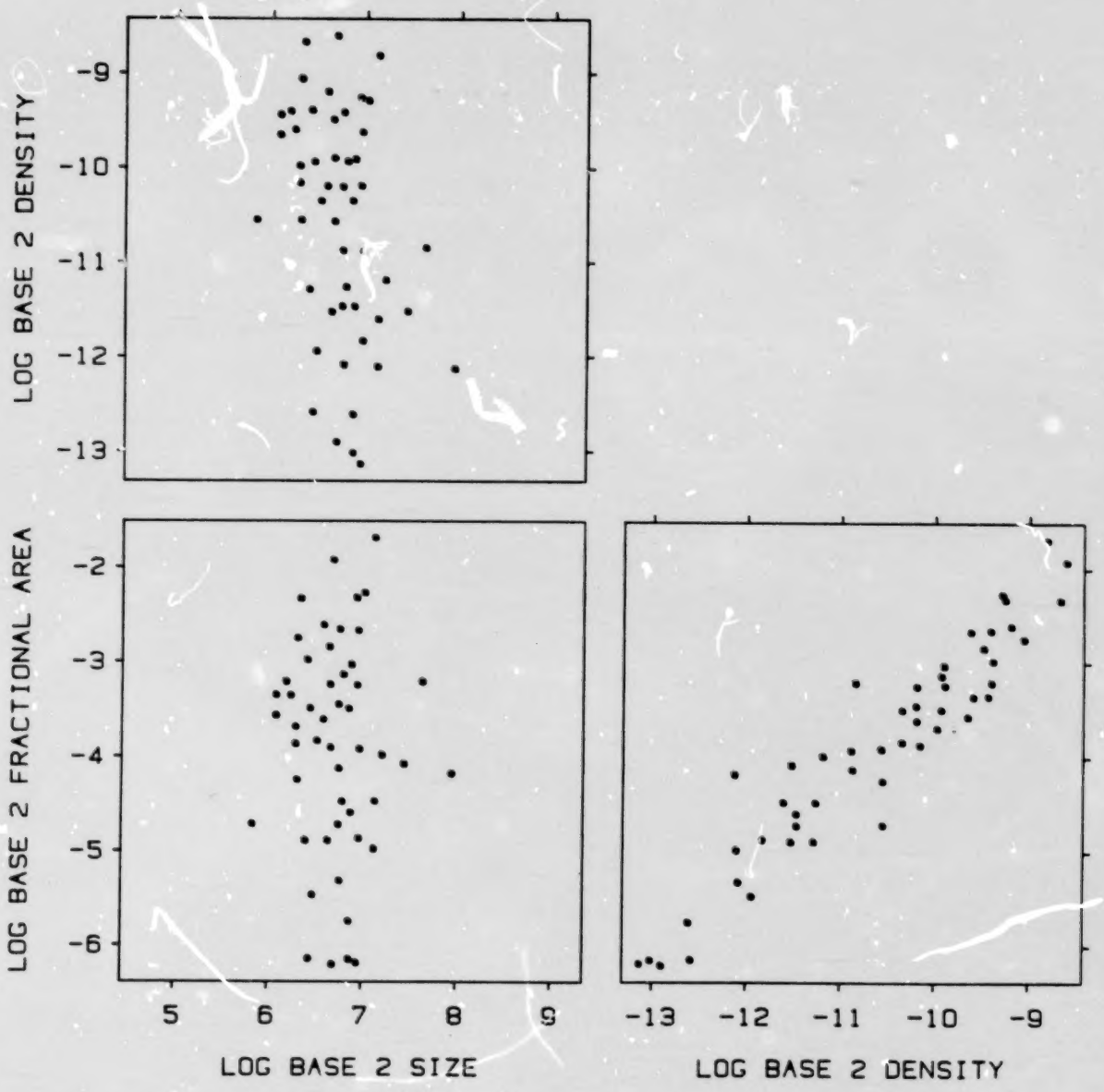

Figure 6. AREA, SIZE, AND DENSITY FOR 50 JOURNALS. Log fractional graph area, log size, and log density for the 50 journals listed in Figure 4 are portrayed. All logarithms are base 2. Density, which is equal to fractional graph area divided by average size, is the average number of graphs per unit area of all articles in a journal. Thus log fractional area equals log size plus log density. Each panel is a scatterplot of two of the three variables. The number of log units per cm on all six axes of the three panels is the same. The figure shows that variation in fractional graph area among the journals is due more to variation in the number of graphs than to variation in size.